



Joint Bayesian Hierarchical Inversion-Classification and Application in Proteomics

Pascal Szacherski, Jean-François Giovannelli, Pierre Grangeat

► To cite this version:

Pascal Szacherski, Jean-François Giovannelli, Pierre Grangeat. Joint Bayesian Hierarchical Inversion-Classification and Application in Proteomics. 2011 IEEE Workshop Statistical Signal Processing, Jun 2011, Nice, France. hal-00585529

HAL Id: hal-00585529

<https://hal.science/hal-00585529>

Submitted on 13 Apr 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

JOINT BAYESIAN HIERARCHICAL INVERSION-CLASSIFICATION AND APPLICATION IN PROTEOMICS

Pascal Szacherski^{1,2}, Jean-François Giovannelli², Pierre Grangeat¹

¹ CEA-LETI, Minatec Campus, 17 rue des Martyrs, 38054 Grenoble Cedex 9, France

² IMS (Université Bordeaux 1 – CNRS – IPB), 351 cours de la Libération, 33405 Talence Cedex, France

ABSTRACT

In this paper, we combine inverse problem and classification for LC-MS data in a joint Bayesian context, given a set of biomarkers and the statistical characteristics of the biological classes. The data acquisition is modelled in a hierarchical way, including random decomposition of proteins into peptides and peptides into ions associated to peaks on the LC-MS measurement. A Bayesian global inversion, based on the hierarchical model for the direct problem, enables to take into account the biological and technological variabilities from those random processes and to estimate the parameters efficiently.

We describe the statistical theoretical framework including the hierarchical direct model, the prior and posterior distributions and the estimators for the involved parameters. We resort to the MCMC algorithm and give preliminary results on a simulated data set.

Index Terms—classification, optimal estimation, quantification, Bayesian inversion, inverse problems, hierarchical model, proteomics, LC-MS

1. INTRODUCTION

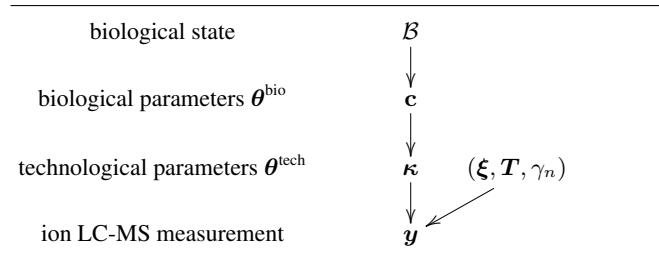
Proteins of an organism are differently expressed according to biological states (healthy, pathological). Those specific proteomic profiles can be used for diagnosis, early detection, therapy planning and follow-up, drug development, etc [1]. Nevertheless, the reconstruction of proteomic profiles remains a challenge due to small and variable concentration of the biomarkers. Furthermore, biomarkers are present within a large protein content including an abundance ratio of up to 10^8 , hence the need to recognise efficiently the biomarkers on molecular profiles.

The reconstruction task has been tackled by several methods: non parametric methods such as area under peak, PLS, N-PLS, PARAFAC [2], parametric methods based either on deterministic least square fitting or on other statistical estimation using for example Bayesian inference [3, 4]. Diagnosis based on proteomic profiles has been shown specific and sensitive [3].

In this communication, our aim is to evaluate the biological state of a sample from a LC-MS (Liquid Chromatography, Mass Spectrometer) measurement. In order to do so, we will develop a classification method in order to optimise diagnoses and their degree of belief.

MS approaches are very attractive to detect protein panels in a sensitive and high speed way. MS can be coupled to many separation techniques. In our development, an LC is used. The physical model of the instrument has been realised by [2]. According to the analytical chain, proteins have to be digested into peptides before introduction into the LC. They will be ionised afterwards to be injected

Fig. 1 – Hierarchical direct model used in our framework for the classification.



in the MS. Both steps are subject to fluctuating behaviour which introduces a technological variability [5].

The biomarker distributions are needed for the classification processing. They can be determined by different procedures: given by an expert, or learnt either on estimations of the biomarker concentrations or on the spectrograms themselves.

In this paper, a hierarchical Bayesian approach to solving the underlying inverse problem is presented. By adopting an inverse problem framework, information about the variables, the instrument, the hierarchical structure, and the molecular interactions are included. The joint posterior distribution for the parameters is expressed, including the hierarchical structure [6, 7]. The global processing makes use of the technological and the biological variabilities which leads to more robust results compared to a succession of step by step processing [8, 9]. This task is not trivial since the proposed approach deals with both continuous and discrete variables.

Classification can be done after estimation of the protein concentration, or directly on the data without estimating other variables. In the proposed method, the originality is to combine inversion and classification by approximating the posterior for all parameters. Finally, estimating the biological state is equivalent to a classification processing.

This document is structured in the following manner. In section 2, the main theory for the development is exposed. Its application within an algorithm is described in section 3. Results are discussed in section 4 before concluding in section 5.

2. STATISTICAL THEORETICAL FRAMEWORK

2.1. Hierarchical direct model of an analytical chain

We are in presence of a biological sample with an unknown biological state. The state is described by the discrete variable \mathcal{B} allowing for two values, healthy H and pathological P . Depending on the biological state, proteins are expressed by their concentration \mathbf{c} . In

order to introduce the sample into the LC-MS, the selected proteins are digested into peptides. The digestion gain is described by a deterministic matrix $\mathbf{D} \in \mathbb{N}^{P \times I}$ translating the number of copies of peptide $i = 1, \dots, I$ produced by protein $p = 1, \dots, P$ [2, ch. 2]. The peptide concentration is denoted κ . None the less, this process is fluctuating, leading to an incomplete digestion. Within the instruments, the peptides are separated by their retention time T in the LC and are subjected to a system gain ξ , characterising in particular the ionisation and the injection within the MS. These steps, in addition to a white Gaussian noise of precision γ_n , lead to the output data \mathbf{y} (see Fig. 1).

The system is calibrated by using the PSAQ standardisation method which adds isotopic standards of the biomarkers in known quantity to the initial sample [10]. These standards are then subjected to the same gains and variabilities than the target proteins, hence they bring information to estimate gains.

2.2. Prior distributions

As shown in Fig. 1, variables are grouped together and referred to as biological and technological parameters, according to the description above. The biological parameter vector θ^{bio} of this framework only contains the concentration \mathbf{c} . The technological parameters are those involved in the instrument, so $\theta^{\text{tech}} = (\kappa, \xi, T, \gamma_n)$.

With respect to the hierarchical structure and to the prior knowledges, we choose a normal distribution for the data \mathbf{y} given the parameters which is also the total likelihood $p(\mathbf{y}|\theta^{\text{tech}}) = \mathbf{N}(\mathbf{y}; \mathbf{H}\kappa, \Gamma_n)$. The system matrix $\mathbf{H} = \mathbf{H}(\xi, T)$ expresses the linearity between the data and the peptide concentration. The diagonal matrix $\Gamma_n = \gamma_n \mathbf{I}$ is the noise precision (inverse covariance) with \mathbf{I} the identity matrix. By doing so, the noise is modelled with the same level on the whole data.

The conditional prior distribution for θ^{tech} is given by the product of the conditional priors for each component. We choose following distributions:

- normal for the gain, $p(\xi) = \mathbf{N}(\xi; \xi_0, \Gamma_\xi)$ with mean ξ_0 and precision Γ_ξ ;
- uniform for the retention time, $p(T) = \mathbf{U}(T^{\min}, T^{\max})$ where T^{\min} is the lower and T^{\max} the upper bound of the allowed interval for T ;
- Gamma for the inverse variance of the noise, $p(\gamma_n) = \mathbf{G}(\gamma_n; \alpha_n, \beta_n)$ where α_n and β_n are the steering parameters of the distribution;
- normal for the peptide concentration given the protein concentration \mathbf{c} , $p(\kappa|\mathbf{c}) = \mathbf{N}(\kappa; \mathbf{D}\mathbf{c}, \Gamma_\kappa)$ with mean $\mathbf{D}\mathbf{c}$ and precision Γ_κ , \mathbf{D} being the digestion matrix (see [2]).

As regards the distribution for θ^{bio} , only the distribution for the protein concentration has to be considered. This distribution is crucial for the classification since it describes the different laws between the healthy ($\mathcal{B} = H$) and the pathological case ($\mathcal{B} = P$). A normal distribution is attributed to protein concentration prior conditionally on the biological state, $p(\mathbf{c}|\mathcal{B}) = \mathbf{N}(\mathbf{c}; \mathbf{m}_\mathcal{B}, \Gamma_\mathcal{B})$ with mean $\mathbf{m}_\mathcal{B}$ and precision $\Gamma_\mathcal{B}$ depending on the biological state \mathcal{B} . That is, by marginalising \mathcal{B} , the distribution for \mathbf{c} is characterised as a Gaussian mixture model.

Finally, to complete the prior, we choose a Bernoulli distribution for the biological state \mathcal{B} , $\Pr(\mathcal{B}) = p_\mathcal{B} \geq 0$ with $p_H + p_P = 1$.

All distributions have been chosen for their good approximation of the real process and for their ability for simple calculations. The

total joint distribution is

$$p(\mathcal{B}, \theta^{\text{bio}}, \theta^{\text{tech}}, \mathbf{y}) = p(\mathbf{y}|\theta^{\text{tech}}) p(\theta^{\text{tech}}|\theta^{\text{bio}}) p(\theta^{\text{bio}}|\mathcal{B}) \Pr(\mathcal{B}). \quad (1)$$

In this work, non informative priors are chosen for all variables except T since the retention time for a peptide is known in a temporal interval. The non-informativity for the variables with normal prior is reached by tending the precisions to zero. For the noise level γ_n , this is done by choosing a Jeffreys prior [6, Appendix A].

2.3. Statistical joint posterior for inversion/classification

We use the joint posterior distribution, $p(\mathcal{B}, \theta^{\text{bio}}, \theta^{\text{tech}}|\mathbf{y})$, given by Bayes' rule as ratio of the joint distribution and the marginalised distribution for \mathbf{y} . In order to classify, the posterior probability of the biological states has to be calculated, i.e. $\Pr(\mathcal{B} = H|\mathbf{y})$ and $\Pr(\mathcal{B} = P|\mathbf{y})$. For this, all parameters except \mathcal{B} are marginalised in the joint posterior:

$$\Pr(\mathcal{B}|\mathbf{y}) = \int p(\mathcal{B}, \theta^{\text{bio}}, \theta^{\text{tech}}|\mathbf{y}) d(\theta^{\text{bio}}, \theta^{\text{tech}}). \quad (2)$$

By integrating them out, all their jointly possible values (biological and technological variability) are considered. Classification is then done by the posterior maximiser of \mathcal{B} , i.e.

$$\hat{\mathcal{B}} = \arg \max_{\mathcal{B} \in \{H, P\}} \Pr(\mathcal{B}|\mathbf{y}). \quad (3)$$

For other decision methods, see [7, sect 5.2].

The posterior mean (PM) given the state $\hat{\mathcal{B}}$ of the other parameters is calculated:

$$[\bar{\theta}_\mathcal{B}^{\text{bio}}, \bar{\theta}_\mathcal{B}^{\text{tech}}] = \int_{\Theta^{\text{bio}} \times \Theta^{\text{tech}}} [\theta^{\text{bio}}, \theta^{\text{tech}}] p(\theta^{\text{bio}}, \theta^{\text{tech}}|\mathbf{y}, \hat{\mathcal{B}}) d(\theta^{\text{bio}}, \theta^{\text{tech}}). \quad (4)$$

This integrates again the variabilities that we have already mentioned and uses the same posterior distribution as in (2). The PM is chosen for the minimisation of the mean quadratic error [7, sect. 2.5].

Solving equations (2) and (4) is analytically impossible. We thus approximate the joint posterior distribution using the MCMC method [6, ch. 11]. This will allow both to marginalise the parameters in equation (2) and to compute the estimator in (4).

3. IMPLEMENTATION

In order to compute the joint posterior by an MCMC method, a Gibbs structure is adopted to transform a global problem into several simpler subproblems. The conditional posteriors for each subproblem depend only on a reduced number of parameters due to the hierarchy [7, sect. 10.2]. Taking into account the conditional independences within each level of hierarchy, the conditional posterior for a parameter is proportional to the product of only its (conditional) prior and the corresponding likelihood (see Sect. 2.2). As an example, the protein concentration is independent from all parameters given the peptide concentration and the biological state: $p(\mathbf{c}|\xi, T, \kappa, \gamma_n, \mathcal{B}) = p(\mathbf{c}|\kappa, \mathcal{B})$. This simplifies the algorithm.

The conditional posteriors for the concentrations (proteins and peptides), the gains and the noise precision are of the same family as their prior respectively, since their priors are conjugated by the corresponding conditional likelihood. Recalling the last example, the conditional posterior for the protein concentration is a Gaussian

Alg. 1 – Pseudo-code of the algorithm used for the Bayesian classification and estimation.

-
1. *initialisation, allocation*
 2. *Gibbs loop*
 for $k = 1$ to $K_0 + K$ do
 - (a) *sample* $\gamma_n^{(k)} \sim p(\gamma_n | \mathbf{y}, \boldsymbol{\kappa}^{(k-1)}, \boldsymbol{\xi}^{(k-1)}, \mathbf{T}^{(k-1)})$
 - (b) *sample* $\boldsymbol{\xi}_n^{(k)} \sim p(\boldsymbol{\xi} | \mathbf{y}, \boldsymbol{\kappa}^{(k-1)}, \gamma_n^{(k)}, \mathbf{T}^{(k-1)})$
 - (c) *sample* $\mathbf{T}^{(k)} \sim p(\mathbf{T} | \mathbf{y}, \boldsymbol{\kappa}^{(k-1)}, \boldsymbol{\xi}^{(k)}, \gamma_n^{(k)})$
 - (d) *sample* $\boldsymbol{\kappa}^{(k)} \sim p(\boldsymbol{\kappa} | \mathbf{y}, \boldsymbol{\xi}^{(k)}, \gamma_n^{(k)}, \mathbf{T}^{(k)})$
 - (e) *sample* $\mathbf{c}^{(k)} \sim p(\mathbf{c} | \boldsymbol{\kappa}^{(k)})$
 - (f) *sample* $\mathcal{B}^{(k)} \sim p(\mathcal{B} | \mathbf{c}^{(k)})$
 3. *approximation of $\Pr(\mathcal{B} | \mathbf{y})$ by histogramming on marginalised samples of $\mathcal{B}^{(k)}$, $k = K_0 + 1, \dots, K_0 + K$, and return $\hat{\mathcal{B}} = \arg \max_{\mathcal{B} \in \{H, P\}} \Pr(\mathcal{B} | \mathbf{y})$.*
 4. *approximation of PM $[\hat{\mathbf{c}}, \hat{\boldsymbol{\kappa}}, \hat{\mathbf{T}}, \hat{\boldsymbol{\xi}}, \hat{\gamma}_n]$ by averaging the samples $[\mathbf{c}^{(k)}, \boldsymbol{\kappa}^{(k)}, \mathbf{T}^{(k)}, \boldsymbol{\xi}^{(k)}, \gamma_n^{(k)}]$ for all $k = K_0 + 1, \dots, K_0 + K$ such that $\mathcal{B}^{(k)} = \hat{\mathcal{B}}$.*
-

distribution with $\mathbf{m}_c^{\text{post}}$ as mean and Γ_c^{post} as precision: $p(\mathbf{c} | \boldsymbol{\kappa}, \mathcal{B}) = \mathbf{N}(\mathbf{c}; \mathbf{m}_c^{\text{post}}, \Gamma_c^{\text{post}})$. This means another simplification: it enables to sample from their conditional posterior easily because they are completely known.

Due to the non informative priors, the conditional posterior distribution for a parameter is determined only by the parameters on its hierarchical level and those below it in Fig. 1, i.e. the level above does not contribute to the conditional posterior here.

For all parameters, sampling is possible. Except for the retention time \mathbf{T} , the prior is conjugated by the likelihood, i.e. the conditional posterior is of the same family as the prior. Within the Gibbs algorithm, this allows for an explicit sampling of these parameters. As to \mathbf{T} , a Metropolis-Hastings Random-Walk step is introduced into the Gibbs loop [6, sect. 11.5]. This is notably a good compromise between an Independent Metropolis-Hastings having fast computation, but slow convergence properties and sophisticated MH methods such as Langevin, Hessian or Hamiltonian updates having slow computation but fast convergence properties [6, ch. 13], [11]. With this method, the sample distribution converges to the posterior distribution, with only little computational effort.

Finally, having simulated enough samples, we approximate the PM and the posterior state probability:

- Probability of the biological state is obtained by marginalisation of all parameters, see Sect. 2.3. Since we sample under the joint posterior, the marginalisation is done by histogramming the samples for \mathcal{B} , regardless the other variables. This empirical probability is used to determine the posterior maximiser MAP, $\hat{\mathcal{B}}$, by the state with the highest empirical probability.
- The conditional posterior mean of the continuous variables is approximated by the empirical average of the samples k such that $\mathcal{B}^{(k)} = \hat{\mathcal{B}}$.

Note that in both cases, we allow for a burn-in period to adjust the distributions since the first samples are not distributed under the posterior [6, sect. 11.6]. We choose arbitrarily to consider only the second half of the samples. The algorithm is summarised in Alg. 1.

4. RESULTS

Testing cohorts of 200 healthy and 200 pathological samples with one biomarker were created, using the instrument model exposed in [4], based upon real data parameters from former studies [3]. In the healthy cohort, the true protein concentration is distributed under $p^*(c|H) = \mathbf{N}(c; m_H^*, \gamma_H^*)$ of mean $m_H^* = 50$ and precision $\gamma_H^* = 0.2$, in the pathological cohort, proteins are distributed under $p^*(c|P) = \mathbf{N}(c; m_P^*, \gamma_P^*)$ of mean $m_P^* = 70$ and precision $\gamma_P^* = 0.2$. These distributions correspond to the biological variability. From one experiment to another, technological variability has been simulated by using different values for θ^{tech} .

The characteristic distributions for each cohort are learnt from 100 learning samples for each biological state using [12]. They are given by $p(c|H) = \mathbf{N}(c; m_H, \gamma_H)$ and $p(c|P) = \mathbf{N}(c; m_P, \gamma_P)$, where $m_H = 50.04$, $m_P = 70.13$, $\gamma_H = 0.12$, $\gamma_P = 0.11$. The learnt Gaussian distributions are wider than the cohort simulation distributions because of the incertitudes in the learning process.

The quantification part of the presented method has been tested successfully on simulated data with one biomarker. Fig. 2(a) demonstrates the dynamic of the sampler. A burn-in period is necessary for the samples being drawn under the posterior distribution. For the sake of convenience, only the marginalised draws for the concentration are shown. After a short period, the concentration samples are drawn under the posterior distribution. In Fig. 2(b), we present the marginalised histogram of the protein concentration from a simulated data with true concentration of 50.07. In contrary to the flat, non informative prior distribution for the concentration, one can see a peaked Gaussian shape with a mean near to the true concentration.

Using the test set, empirical probabilities have been approximated. In Fig. 2(c), a histogram showing the degree of belief in $\hat{\mathcal{B}}_n = H$ is presented. Firstly, the figure describes that *all* tested data have been classified correctly since the classification is done upon the empirical posterior probability for the biological states. The false detection rate, i.e. $\Pr(\hat{\mathcal{B}} = B | \mathcal{B} = B^*)$, with $B \neq B^*$, is zero. Secondly, the degree of belief expresses the method's certitude in its estimation, given by the empirical probability for the estimated biological state. In the given configuration, the certitude in the classification is of 96% or higher which reflects as well the probability of membership to the estimated biological state. In other words, for nearly all data, the probability that biological state given the data is the estimated one is 100%, so one can be sure to classify correctly.

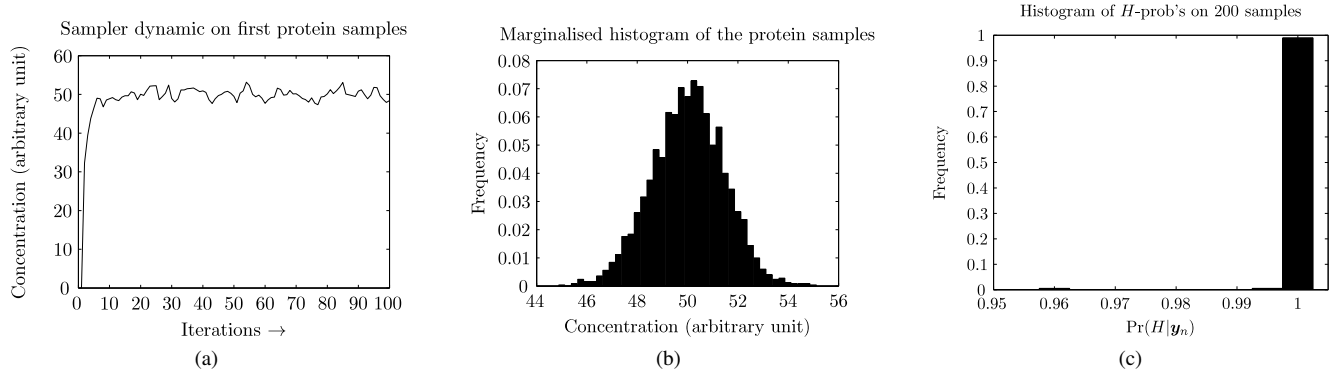
The considered configuration favours good estimations since the classes are well separated. As the mean m_P of the pathological class approaches the mean m_H of the healthy class, the classification performance naturally decreases. The distance of the two considered classes can be measured by the Kullback-Leibler-divergence. We define de *Right Decision Rate* (RDR) as the ratio of number of right estimations over the total number of estimations, and the *False Decision Rate* (FDR) as the ratio of number of false estimations over the total number of estimations, corresponding to a Type-I-error.

The results of the performances are shown in Tab. 1. For KL greater than 6.83, no false estimation is done, that is the FDR is 0. For common values of the FDR, i.e. 5% and 10%, the classification with the presented method is still done satisfactorily with class distributions that are near and, thus, overlapping each other (KL ≈ 4.60 and KL ≈ 2.79 respectively).

5. CONCLUSION

In this paper, we presented an original way of combining inverse problem and classification in proteomics by using an indirect mea-

Fig. 2 – (a) Dynamic of the Gibbs sampler on the first 100 iterations. After only a few iterations, the samples gives random values from the posterior distribution. (b) Histogram of 2500 marginalised samples for the protein concentration. One can see the shape of a Gaussian function with mean $m = 50.012$ (real concentration 50.07) and precision $\gamma = 0.48$. (c) Histogram of approximated probabilities $\Pr(H|y_n)$ on 200 simulated test samples. $\Pr(H|y_n) \geq 0.96$ for $n = 1, \dots, 100$.



Tab. 1 – Robustness test: classification performance when m_P tends to $m_H = 50.04$, given the the precisions $\gamma^H = 0.12$ and $\gamma^P = 0.11$. RDR stands for the right decision rate, FDR for false decision rate; KL is the Kullback-Leibler-divergence.

m_P	KL	RDR	FDR
70.13	22.4	1.00	0.00
61.13	6.83	1.00	0.00
59.13	4.60	0.96	0.04
57.13	2.79	0.91	0.09
54.13	0.931	0.80	0.20
52.13	0.244	0.66	0.34
50.04	0	0.50	0.50

surement of the classifying variable in a LC-MS spectrum. We have shown that this approach yields good results which are robust due to the probabilistic direct modelling of each biological and technical step and the corresponding statistical inversion.

This framework improves both the classification and the quantification when the prior on the concentration is conditioned by the class of the samples. This will allow to combine within the same statistical framework quantitative and differential proteomic analysis.

Later on, the multivariate approach will allow to describe possible interactions between proteins. The work is transposable to any comparison or differential analysis, other than diagnostics.

We restricted this work to a binary classification, but a classification with more classes, such as different pathologies or different states of the same pathology, can also be performed.

6. REFERENCES

- [1] M. Palmblad, A. Tiss, and R. Cramer, “Mass spectrometry in clinical proteomics - from the present to the future,” *Proteomics – Clinical Applications*, vol. 3, no. 1, pp. 6–17, 2009.
- [2] G. Strubel, *Reconstruction de profils moléculaires: modélisation et inversion d’une chaîne de mesure protéomique*, Ph.D. thesis, École Polytechnique de Grenoble, France, 2008.
- [3] P. Grangeat et al., “First demonstration on NSE biomarker of a computational environment dedicated to lab-on-chip based cancer diagnosis,” Poster at 58th ASMS Conference, Salt Lake City, USA, 2010.
- [4] G. Strubel, J.-F. Giovannelli, C. Paulus, L. Gerfault, and P. Grangeat, “Bayesian estimation for molecular profile reconstruction in proteomics based on liquid chromatography and mass spectrometry,” in *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, France, pp. 5979–5982, 2007.
- [5] B. Lumbreras et al., “Sources of error and its control in studies on the diagnostic accuracy of -omics technologies,” *Proteomics – Clinical Applications*, vol. 3, no. 2, pp. 173–184, 2009.
- [6] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis, Second Edition (Texts in Statistical Science)*, Chapman & Hall/CRC, 2 edition, July 2003.
- [7] C. Robert, *The Bayesian choice : from decision-theoretic foundations to computational implementation*, Springer, New York, NY, 2007.
- [8] L. Gerfault, G. Strubel, C. Paulus, J.-F. Giovannelli, and P. Grangeat, “évaluation statistique d’un algorithme bayésien pour la reconstruction de profils moléculaires par spectrométrie de masse,” in *XXII^{ème} Colloque GRETSI*, Dijon, France, 2009.
- [9] P. Grangeat, G. Strubel, J. F. Giovannelli, V. Brun, L. Gerfault, C. Paulus, A. Dupuis, and J. Garin, “Robust statistical reconstruction of protein profiles in mass spectrometry,” Poster at 57th ASMS Conference, Philadelphia, USA, 2009.
- [10] V. Brun, C. Masselon, J. Garin, and A. Dupuis, “Isotope dilution strategies for absolute quantitative proteomics,” *Journal of Proteomics*, vol. 72, no. 5, pp. 740–749, 2009.
- [11] C. Vacar, J.-F. Giovannelli, and Y. Berthoumieu, “Langevin and Hessian with Fisher approximation: Stochastic sampling for parameter estimation of structured covariance,” in *ICASSP*, Prague, Czech Republic, May 2011.
- [12] P. Szacherski, J.-F. Giovannelli, L. Gerfault, and P. Grangeat, “Apprentissage supervisé robuste de caractéristiques de classes. Application en protéomique,” *submitted to XXIII^{ème} colloque GRETSI*, 2011.